CLAIMS

1. A method of determining regression functions from a computer data input, the regression functions for use in data mining, prediction, calibration, segmentation or response analysis, the method using K-Harmonic Means regression clustering and comprising the steps of:

selecting K regression functions $f_1,\ldots,f_K$;

associating an i-th data point from a dataset with a k-th regression function using a soft membership function;

providing a weighting to each data point using a weighting function to determine a particular data point's participation in a calculation of a residue error;

calculating a residue error between a weighted i-th data point and its associated regression function;

iterating to minimize a total residue error; and

identifying suitable regression functions for use in the analysis.

2. The method of claim 1, wherein the soft membership function provides a probabilistic association between the i-th data point and the k-th regression function.

3. The method of claim 1, wherein not all data points of the computer data input fully participate in each iteration.

4. The method of claim 1, wherein the total residue error is described by a function in L^q - space and the parameter q is greater than 2.

5. A method of determining regression functions from a computer data input $Z=(X,Y)=\{(x_i,y_i)\mid i=1,\ldots,N\}$, the regression

segmentation or response analysis, the method using *K*-Harmonic Means regression clustering and comprising the steps of:

selecting *K* regression functions $f_1,...,f_K$, in an r-th iteration;

associating an i-th data point from the dataset Z with a k-th regression function f_k using a soft probability membership function that can be expressed as,

$p(Z_k \mid z_i) = d_{i,k}^{p+q} \Big/ \sum_{l=1}^{K} d_{i,l}^{p+q}$ where, $d_{i,k} = \| f_k^{(r-1)}(x_i) - y_i \|$, $p \geq 2$, and where q is a variable parameter;

providing a weighting to each data point z_i using a weighting function that can be expressed as,

$$a_p(z_i) = \sum_{l=1}^{K} d_{i,l}^{p+q} \Big/ \sum_{l=1}^{K} d_{i,l}^{p}$$

to determine the data point's participation in calculating a residue error;

calculating a residue error between a weighted i-th data point and its associated regression function;

iterating to minimize a total residue error; and

identifying suitable regression functions for use in the analysis.

6. The method of claim 5, wherein the residue error is defined by an error function that can be expressed as $e(f_k(x_i), y_i) = \| f_k(x_i) - y_i \|^p$, with $p \geq 2$.

7. The method of claim 5, wherein a regression optimization is used that satisfies,

$$f_k^{(r)} = \arg\min_{f \in \Phi} \sum_{i=1}^{N} a_p(z_i) p(Z_k \mid z_i) \| f(x_i) - y_i \|^q$$, for $k = 1,...,K$.

8. The method of claim 5, wherein the K-Harmonic Means objective function is defined by replacing the MIN() function of a K Means objective function with a harmonic average, HA(), function.

9. The method of claim 5, wherein iterations are done until the change in the objective function between iterations is less than a predetermined threshold.

10. The method of claim 8, wherein the objective function can be expressed as:

$$Perf_{RC-KHM_p}(Z,M)=\sum_{i=1}^{N}\underset{1\le k\le K}{HA}\{\| f_k(x_i)-y_i\|^p\}=\sum_{i=1}^{N}\frac{K}{\sum_{k=1}^{K}\frac{1}{\| f_k(x_i)-y_i\|^p}}.$$

11. The method of claim 5, wherein the determined regression functions are used to predict outcomes from a new dataset.

12. The method of claim 11, wherein the predictions from the K regression functions on a new data point are provided as K possible values each with an associated probability.

13. The method of claim 5, wherein the determined regression functions are additionally used to initialize an Expectation Maximization regression clustering method.

14. A system for determining regression functions from a computer data input $Z=(X,Y)=\{(x_i,y_i)\mid i=1,\ldots,N\}$, the system using K-Harmonic Means regression clustering and comprising:

data input and storage means to receive and store the computer data input;

a determined-regression-function display;

a processor providing for:

selecting K regression functions $f_1,\ldots,f_K$, in an r-th iteration;

associating an i-th data point from a dataset Z with a k-th regression function f_k using a soft membership function;

providing a weighting to each data point z_i using a weighting function to determine the data point's participation in calculating a residue error;

calculating a residue error between a weighted i-th data point and its associated regression function;

iterating to minimize a total residue error; and

determining suitable regression functions for output.

15. The system of claim 14, wherein the system calls a regression analysis step as a subroutine from an existing program or library.

16. The system of claim 14, wherein the soft membership function provides a probabilistic association between the i-th data point and the k-th regression function.

17. The system of claim 16, wherein the soft membership function can be expressed as,

$p(Z_k \mid z_i) = d_{i,k}^{p+q} \Big/ \sum_{l=1}^{K} d_{i,l}^{p+q}$ where, $d_{i,k} = \| f_k^{(r-1)}(x_i) - y_i \|$, $p \geq 2$, and where q is a variable parameter.

18. The system of claim 14, wherein the weighting function can be expressed mathematically as,

$$a_p(z_i) = \sum_{l=1}^{K} d_{i,l}^{p+q} \Big/ \sum_{l=1}^{K} d_{i,l}^{p} .$$

19. The system of claim 14, wherein the determined regression functions approximate sales or marketing data, provide economic demand curves, identify segmentation of customer responses, provide calibration parameters, or provide segmentation or interpolation of static or video images.